Farmers Market Analysis

Vinit Deshbhratar

01/02/2020

The aim of the project is to identify trends in the farmer market. For example is it seasonal, type the food category served in the market, the types of transaction methods supported in the farmers market and in the end come up with a derived information which may not be possible to observe directly. The dataset has been taken from the USDA website. The link for the dataset can be found here.

Loading the dataset

```
library(dplyr)
library(lubridate)
library(ggplot2)
library(stringr)
library(tidyr)
```

Importing the dataset

```
dataset = read.csv('farmers_market_info.csv')
```

Taking a look at the dataset

```
head(dataset)
```

```
##      FMID                                        MarketName
## 1 1018261  Caledonia Farmers Market Association - Danville
## 2 1018318                Stearns Homestead Farmers' Market
## 3 1009364                106 S. Main Street Farmers Market
## 4 1010691              10th Steet Community Farmers Market
## 5 1002454                              112st Madison Avenue
## 6 1011100                           12 South Farmers Market
##                                                 Website
## 1 https://sites.google.com/site/caledoniafarmersmarket/
## 2                       http://www.StearnsHomestead.com
## 3               http://thetownofsixmile.wordpress.com/
## 4
## 5
## 6                  http://www.12southfarmersmarket.com
##                                             Facebook         Twitter
## 1 https://www.facebook.com/Danville.VT.Farmers.Market/
## 2                       StearnsHomesteadFarmersMarket
## 3
## 4
## 5
## 6                             12_South_Farmers_Market @12southfrmsmkt
```

```
##   Youtube
## 1
## 2
## 3
## 4
## 5
## 6
##                                                          OtherMedia
## 1
## 2
## 3
## 4 http://agrimissouri.com/mo-grown/grodetail.php?type=mo-grown&ID=275
## 5
## 6                                                    @12southfrmsmkt
##                   street      city    County          State   zip
## 1                         Danville Caledonia        Vermont  5828
## 2        6975 Ridge Road    Parma   Cuyahoga           Ohio
## 3     106 S. Main Street  Six Mile           South Carolina 29682
## 4 10th Street and Poplar    Lamar     Barton       Missouri 64759
## 5   112th Madison Avenue  New York  New York       New York 10029
## 6 3000 Granny White Pike Nashville  Davidson      Tennessee 37204
##               Season1Date                                 Season1Time
## 1 06/14/2017 to 08/30/2017                         Wed: 9:00 AM-1:00 PM;
## 2 06/24/2017 to 09/30/2017                         Sat: 9:00 AM-1:00 PM;
## 3
## 4 04/02/2014 to 11/30/2014   Wed: 3:00 PM-6:00 PM;Sat: 8:00 AM-1:00 PM;
## 5         July to November Tue:8:00 am - 5:00 pm;Sat:8:00 am - 8:00 pm;
## 6 05/05/2015 to 10/27/2015                         Tue: 3:30 PM-6:30 PM;
##               Season2Date           Season2Time Season3Date Season3Time
## 1 09/06/2017 to 10/18/2017 Wed: 2:00 PM-6:00 PM;
## 2
## 3
## 4
## 5
## 6
##   Season4Date Season4Time         x        y                     Location
## 1                         -72.14034 44.41104
## 2                         -81.73394 41.37480
## 3                         -82.81870 34.80420
## 4                         -94.27462 37.49563
## 5                         -73.94930 40.79390 Private business parking lot
## 6                         -86.79071 36.11837
##   Credit WIC WICcash SFMNP SNAP Organic Bakedgoods Cheese Crafts Flowers
## 1      Y   Y       N     Y    N       Y          Y      Y      Y       Y
## 2      Y   N       N     Y    N       -          Y      N      Y       Y
## 3      Y   N       N     N    N       -
## 4      Y   N       N     N    N       -          Y      N      Y       N
## 5      N   N       Y     Y    N       -          Y      N      Y       Y
## 6      Y   N       N     N    Y       Y          Y      Y      N       Y
##   Eggs Seafood Herbs Vegetables Honey Jams Maple Meat Nursery Nuts Plants
## 1    Y       N     Y          Y     Y    Y     Y    Y       N    N      N
## 2    Y       N     Y          Y     Y    Y     Y    N       N    N      N
## 3
## 4    Y       N     Y          Y     Y    Y     N    Y       N    N      Y
```

```
## 5    N       N     Y          Y     Y    Y     N    N       N    Y      N
## 6    Y       N     Y          Y     Y    Y     Y    Y       N    N      N
##   Poultry Prepared Soap Trees Wine Coffee Beans Fruits Grains Juices
## 1       Y        Y    Y     Y    N      Y     Y      Y      N      N
## 2       Y        N    Y     N    N      N     N      Y      N      N
## 3
## 4       Y        Y    Y     N    N      N     N      Y      N      N
## 5       N        Y    Y     N    N      N     N      N      N      N
## 6       Y        Y    Y     N    N      Y     N      Y      N      Y
##   Mushrooms PetFood Tofu WildHarvested            updateTime
## 1         Y       Y    N             N 6/20/2017 10:43:57 PM
## 2         N       N    N             N  6/21/2017 5:15:01 PM
## 3                                                      2013
## 4         N       N    N             N 10/28/2014 9:49:46 AM
## 5         N       N    N             N      03-01-2012 10:38
## 6         Y       Y    N             N      05-01-2015 10:40
```

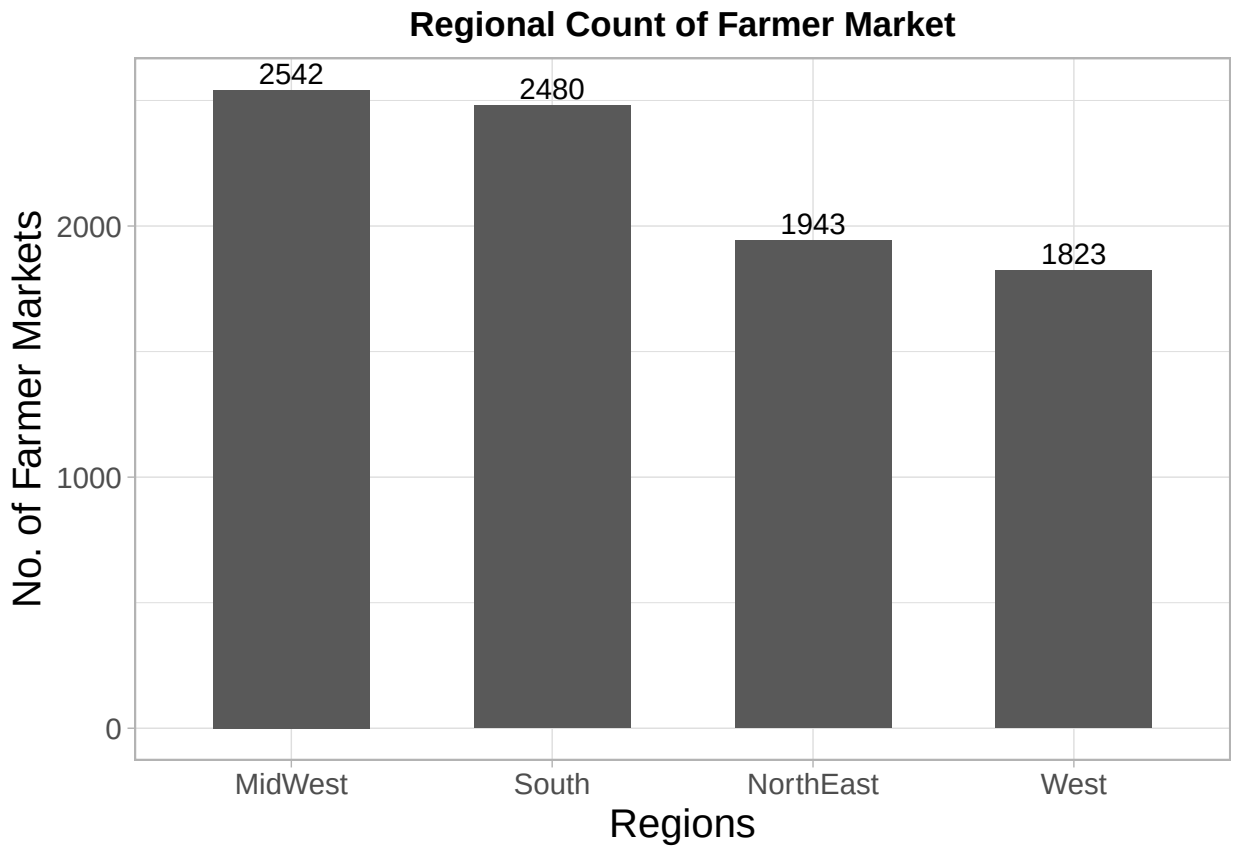
Identify the number of farmer markets

For this task, we'll segregate the states in the country in regions based on geography like, east, west, etc. We'll take a aggregated count of the regions to get the count of farmer markets in each region.

```
# Declaring the states according to the Region
NorthEast = c('Maine', 'New Hampshire', 'Vermont', 'Massachusetts', 'Rhode Island',
              'Connecticut', 'New York', 'New Jersey', 'Pennsylvania')
MidWest = c('Ohio', 'Michigan', 'Indiana', 'Wisconsin', 'Illinois', 'Minnesota',
            'Iowa', 'Missouri', 'North Dakota', 'South Dakota', 'Nebraska', 'Kansas')
South = c('Delaware', 'Maryland', 'Virginia', 'West Virginia', 'Kentucky', 'North Carolina',
          'South Carolina', 'Tennessee', 'Georgia', 'Florida', 'Alabama', 'Mississippi',
          'Arkansas', 'Louisiana', 'Texas', 'Oklahoma', 'Puerto Rico', 'Virgin Islands', 'District of C
West = c('Montana', 'Idaho', 'Wyoming', 'Colorado', 'New Mexico', 'Arizona', 'Utah', 'Nevada',
         'California', 'Oregon', 'Washington', 'Alaska', 'Hawaii')

# Adding a column to the dataset having an additional regional information
dataset = dataset %>%
             mutate(Region = ifelse(State %in% NorthEast, 'NorthEast',
                                    ifelse(State %in% MidWest, 'MidWest',
                                    ifelse(State %in% South, 'South', 'West'))))

regional_dataset = dataset %>%
                      group_by(Region) %>%
                      summarise(Count = n())

# Plotting the dataset
ggplot(data = regional_dataset) +
  geom_histogram(aes(x = reorder(Region, -Count), y = Count), width = 0.6, stat = 'identity') +
  xlab('Regions') +
  ylab('No. of Farmer Markets') +
  theme_light() +
  geom_text(aes(x = Region, y = Count, label = Count, vjust = -0.3), size = 4)+
  ggtitle("Regional Count of Farmer Market") +
  theme(plot.title = element_text(hjust = 0.5, face = 'bold'),
        axis.text = element_text(size = 11),
```

```
        axis.title = element_text(size = 15))
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```



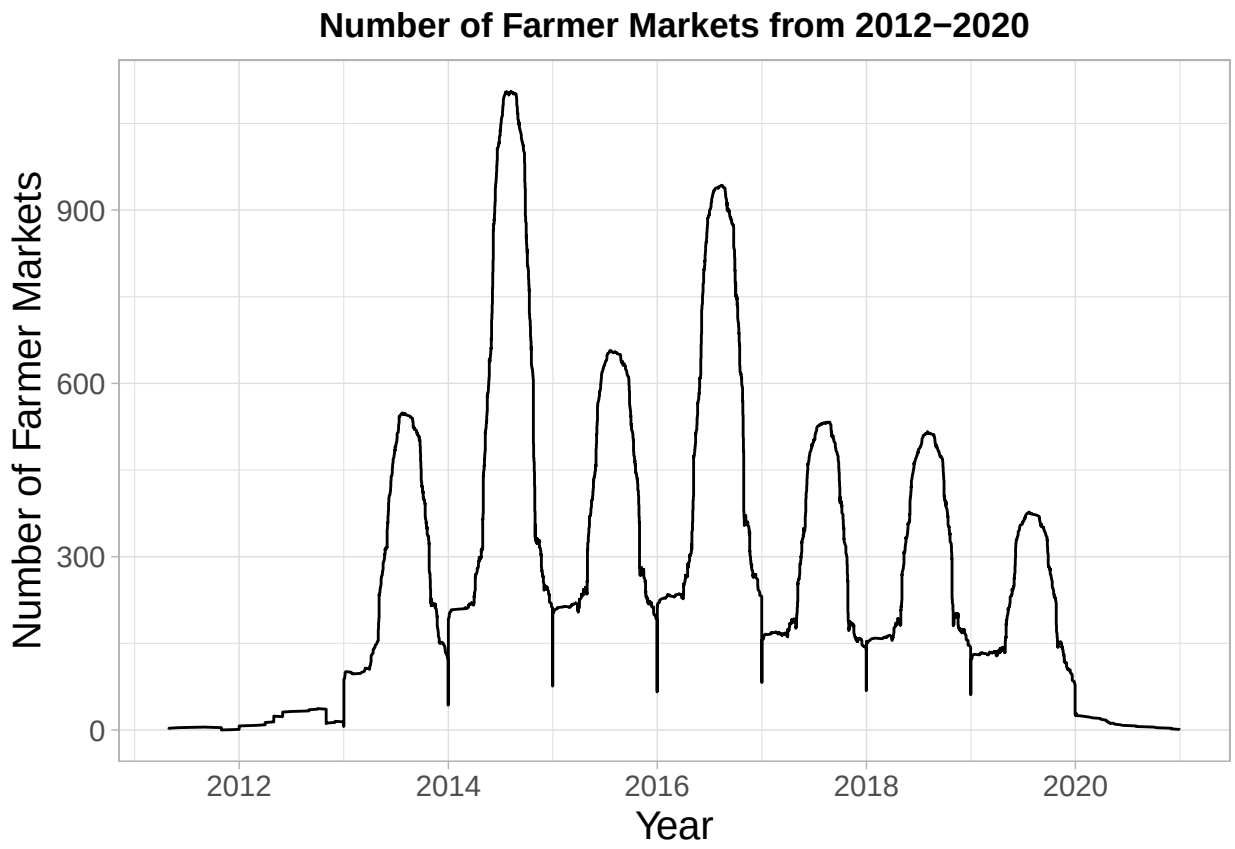
Removing the varibles

```
rm(regional_dataset, MidWest, NorthEast, South, West)
```

Time series analysis of the farmers market

```
task_2 = select(dataset, Region, Season1Date, Season2Date, Season3Date, Season4Date)

task_2$Season1StartDate = str_split(task_2$Season1Date, ' to ', simplify = TRUE)[, 1]
task_2$Season1EndDate = str_split(task_2$Season1Date, 'to ', simplify = TRUE)[, 2]

task_2$Season2StartDate = str_split(task_2$Season2Date, ' to ', simplify = TRUE)[, 1]
task_2$Season2EndDate = str_split(task_2$Season2Date, ' to ', simplify = TRUE)[, 2]

task_2$Season3StartDate = str_split(task_2$Season3Date, ' to ', simplify = TRUE)[, 1]
task_2$Season3EndDate = str_split(task_2$Season3Date, ' to ', simplify = TRUE)[, 2]

task_2$Season4StartDate = str_split(task_2$Season4Date, ' to ', simplify = TRUE)[, 1]
task_2$Season4EndDate = str_split(task_2$Season4Date, ' to ', simplify = TRUE)[, 2]
```

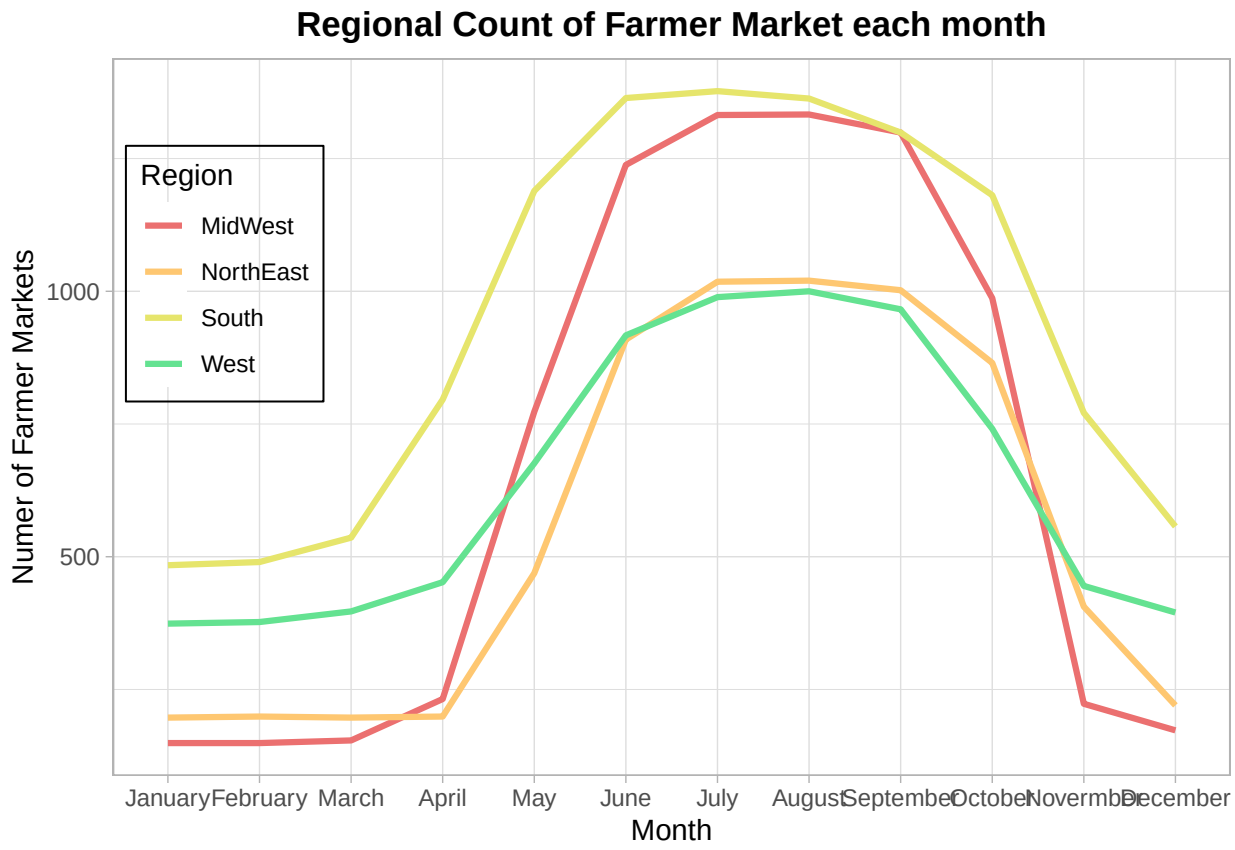
```
task_2 = select(task_2, Region, Season1StartDate, Season1EndDate, Season2StartDate, Season2EndDate
                , Season3StartDate, Season3EndDate, Season4StartDate, Season4EndDate)

season1_df = task_2 %>%
  filter(!is.na(mdy(Season1StartDate)) & !is.na(mdy(Season1EndDate))) %>%
  mutate(StartDate = Season1StartDate) %>%
  mutate(EndDate = Season1EndDate) %>%
  select(Region, StartDate, EndDate)

season2_df = task_2 %>%
  filter(!is.na(mdy(Season2StartDate)) & !is.na(mdy(Season2EndDate))) %>%
  mutate(StartDate = Season2StartDate) %>%
  mutate(EndDate = Season2EndDate) %>%
  select(Region, StartDate, EndDate)

season3_df = task_2 %>%
  filter(!is.na(mdy(Season3StartDate)) & !is.na(mdy(Season3EndDate)))%>%
  mutate(StartDate = Season3StartDate) %>%
  mutate(EndDate = Season3EndDate) %>%
  select(Region, StartDate, EndDate)
season4_df = task_2 %>%
  filter(!is.na(mdy(Season4StartDate)) & !is.na(mdy(Season4EndDate)))%>%
  mutate(StartDate = Season4StartDate) %>%
  mutate(EndDate = Season4EndDate) %>%
  select(Region, StartDate, EndDate)
all_season_df = rbind(season1_df, season2_df, season3_df, season4_df)

all_season_start_df = all_season_df %>%
  mutate(Date = StartDate) %>%
  mutate(Start_End = 1) %>%
  select(Region, Date, Start_End)

all_season_end_df = all_season_df %>%
  mutate(Date = EndDate) %>%
  mutate(Start_End = -1) %>%
  select(Region, Date, Start_End)
all_season_same_df = rbind(all_season_start_df, all_season_end_df)
all_season_same_df$Date = as.Date(all_season_same_df$Date, format = '%m/%d/%Y')
all_season_same_df = all_season_same_df[order(all_season_same_df$Date), ]
all_season_same_df = all_season_same_df %>%
  filter(!is.na(Date))
all_season_same_df$Sum = cumsum(all_season_same_df$Start_End)

ggplot(all_season_same_df) +
  geom_line(aes(x = Date, y = Sum)) +
  xlab('Year') +
  ylab('Number of Farmer Markets') +
  ggtitle('Number of Farmer Markets from 2012-2020') +
  theme_light() +
  theme(plot.title = element_text(hjust = 0.5, face = 'bold'),
        axis.text = element_text(size = 11),
        axis.title = element_text(size = 15))
```



The plot clearly shows the seasonal nature of the farmers market. The most number of markets were in the year 2014 and then it has gone down always except in the year 2016. We'll take a look at the monthly trend of the farmers market.

```
all_season_df$StartDate = month(mdy(all_season_df$StartDate))
all_season_df$EndDate = month(mdy(all_season_df$EndDate))
months = c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12)
months_text = c('January', 'February', 'March', 'April', 'May', 'June', 'July', 'August',
                'September', 'October', 'Novermber', 'December')
all_season_df = all_season_df %>%
  group_by(StartDate, EndDate) %>%
  mutate(Months = ifelse(StartDate < EndDate,
                         paste(months[StartDate:EndDate], collapse = ','),
                         paste(months[-((StartDate-1):(EndDate+1))], collapse = ',')))

all_season_df = separate_rows(all_season_df, Months, sep = ',')

month_df = all_season_df %>%
  group_by(Region, Months) %>%
  summarise(Count = n())
month_df$Months = factor(month_df$Months, levels = months, labels = months_text)

ggplot(month_df) +
  geom_line(aes(x = Months, y = Count, color = Region, group = Region), size = 1.1) +
  theme_light() +
  xlab('Month') +
```

```
  ylab('Numer of Farmer Markets') +
  ggtitle('Regional Count of Farmer Market each month') +
  scale_color_manual(values = c('#EB7070', '#FEC771', '#E6E56C', '#64E291')) +
  theme(plot.title = element_text(hjust = 0.5, face = 'bold'),legend.position = c(0.1, 0.7),
        legend.background = element_blank(), legend.box.background = element_rect(color = 'black'))
```



Midwest and South had almost the same number of farmer markets, but the seasonal nature of the markets is more visible in Midwest, where the number of farmer markets is the least in the starting few months, but it sees a great rise in the number of markets in the month April to July and then goes down from September.

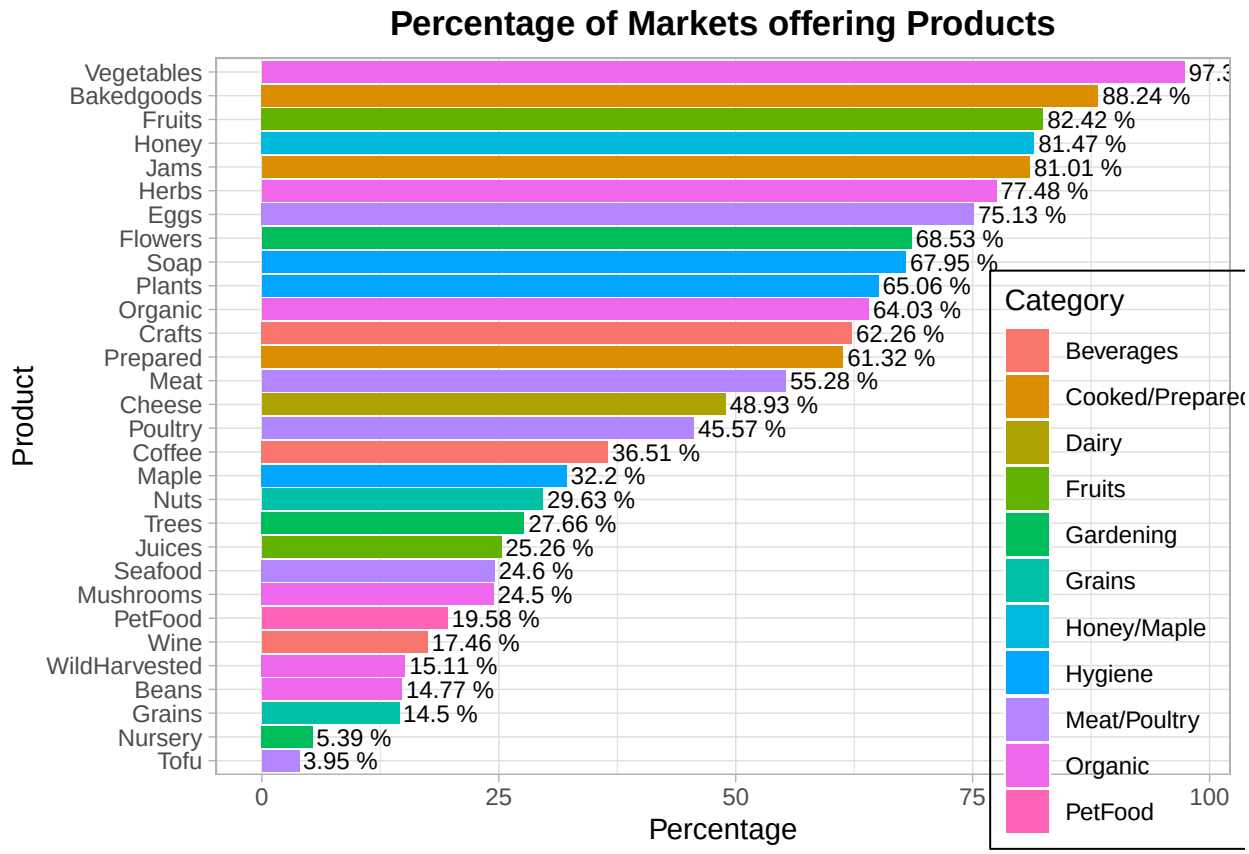
Removing the variables.

```
rm(all_season_df, all_season_end_df, all_season_same_df, all_season_start_df, month_df,
   season1_df, season2_df, season3_df, season4_df, task_2)
```

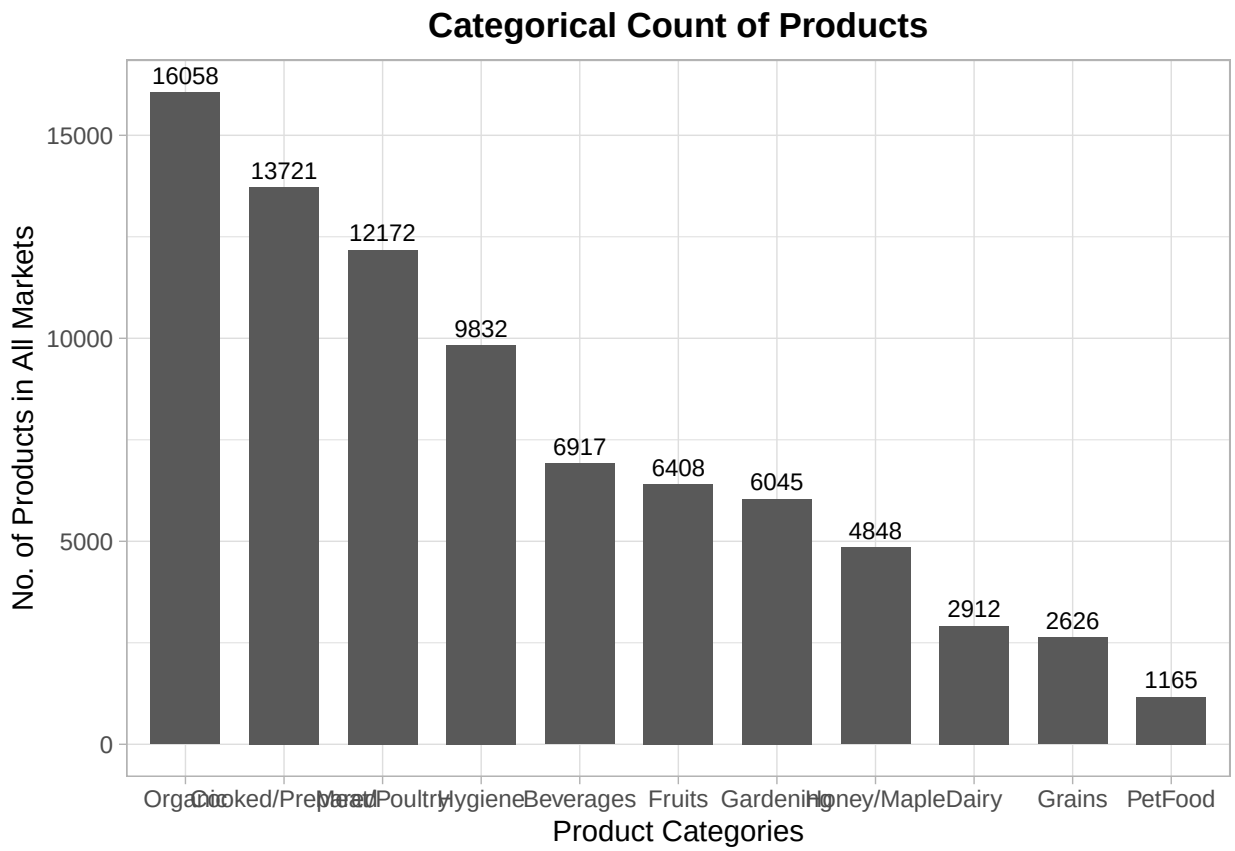
Analysis of the products being offered in the farmers market

```
dairy = c('Cheese')
meat_products = c('Eggs', 'Seafood', 'Meat', 'Poultry', 'Tofu')
added_sugar = c('Honey', 'Mapel')
fruits = c('Fruits', 'Juices')
organic = c('Organic', 'Herbs', 'Vegetables', 'Beans', 'WildHarvested', 'Mushrooms')
grains = c('Nuts', 'Grains')
cooked = c('Bakedgoods', 'Prepared', 'Jams')
```

```
gardening = c('Nursery', 'Trees', 'Flowers')
petfood = c('PetFood')
beverage = c('Crafts', 'Wine', 'Coffee')
hygiene = c('Soap')

col_start_ind = which(colnames(dataset) == 'Organic')
col_end_ind = which(colnames(dataset) == 'WildHarvested')
for(i in col_start_ind:col_end_ind){
  dataset[, i] = factor(dataset[, i], exclude = c('-', ''))
}
count = c()
total_count = c()
for(i in col_start_ind:col_end_ind){
  count = c(count, summary(dataset[, i])['Y'])
  sum = summary(dataset[, i])['Y'] + summary(dataset[, i])['N']
  total_count = c(total_count, sum)
}
col_names = colnames(dataset)[col_start_ind:col_end_ind]
food_df = data.frame(col_names, total_count, count)
colnames(food_df) = c('Product', 'Total_Count', 'Count')

food_df = food_df %>%
  mutate(Category = ifelse(Product %in% dairy, 'Dairy',
                           ifelse(Product %in% meat_products, 'Meat/Poultry',
                                  ifelse(Product %in% added_sugar, 'Honey/Maple',
                                         ifelse(Product %in% fruits, 'Fruits',
                                                ifelse(Product %in% organic, 'Organic',
                                                       ifelse(Product %in% grains, 'Grains',
                                                              ifelse(Product %in% cooked, 'Cooked/Prepar
                                                                     ifelse(Product %in% gardening, 'Gar
                                                                            ifelse(Product %in% petfood,
                                                                                   ifelse(Product %in% b
                                                                                          'Hygiene')))))


food_df = food_df %>%
  group_by(Category) %>%
  mutate(Percentage = round(Count/Total_Count*100, 2))


category_df = food_df %>%
  group_by(Category) %>%
  summarise(Count = sum(Count))

ggplot(data = food_df) +
  geom_bar(aes(x = reorder(Product, Percentage), y = Percentage, fill = Category), stat = 'identity') +
  theme_light() +
  coord_flip() +
  geom_text(aes(x = Product, y = Percentage, label = paste(Percentage, '%')), hjust = -0.05, size = 3) +
  xlab('Product') +
  ggtitle('Percentage of Markets offering Products') +
  theme(plot.title = element_text(hjust = 0.5, face = 'bold'), legend.position = c(0.9, 0.3),
        legend.background = element_blank(), legend.box.background = element_rect(color = 'black'))
```




```
ggplot(data = category_df) +
  geom_histogram(aes(x = reorder(Category, -Count), y = Count, width=0.7), stat = 'identity') +
  geom_text(aes(x = Category, y = Count, label = Count), vjust = -0.5, size = 3) +
  xlab('Product Categories') +
  ylab('No. of Products in All Markets') +
  ggtitle('Categorical Count of Products') +
  theme_light() +
  theme(plot.title = element_text(hjust = 0.5, face = 'bold'))
```



Removing the variables

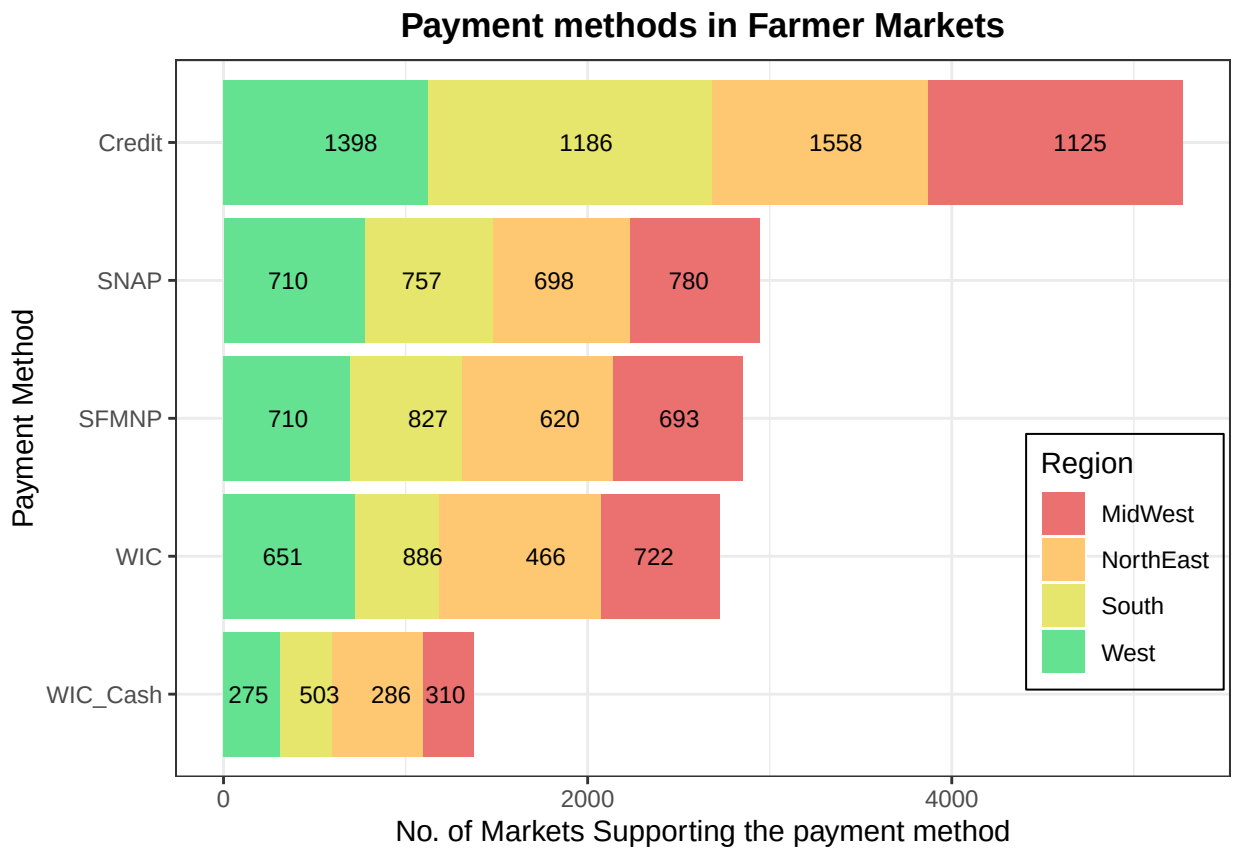
```
rm(category_df, food_df, added_sugar, beverage, col_end_ind, col_start_ind, col_names, cooked, count, da
```

Analysis of the payment methods offered

```
credit_count =  dataset %>%
  group_by(Region) %>%
  filter(Credit == 'Y') %>%
  summarise(Count = n()) %>%
  mutate(Payment_Method = 'Credit')

wic_count = dataset %>%
  group_by(Region) %>%
  filter(WIC == 'Y') %>%
  summarise(Count = n()) %>%
  mutate(Payment_Method = 'WIC')

wiccash_count = dataset %>%
  group_by(Region) %>%
  filter(WICcash == 'Y') %>%
  summarise(Count = n()) %>%
  mutate(Payment_Method = 'WIC_Cash')
```

```r
sfmnp_count = dataset %>%
  group_by(Region) %>%
  filter(SFMNP == 'Y') %>%
  summarise(Count = n()) %>%
  mutate(Payment_Method = 'SFMNP')

snap_count = dataset %>%
  group_by(Region) %>%
  filter(SNAP == 'Y') %>%
  summarise(Count = n()) %>%
  mutate(Payment_Method = 'SNAP')

task3_df = rbind(credit_count, wic_count, wiccash_count, sfmnp_count, snap_count)

ggplot(data = task3_df, aes(x = reorder(Payment_Method, Count), y = Count)) +
  geom_histogram(aes(fill = Region), stat = 'identity') +
  theme_bw() +
  xlab('Payment Method') +
  ylab('No. of Markets Supporting the payment method') +
  geom_text(aes(label = Count), position = position_stack(vjust = 0.5), size = 3) +
  theme(plot.title = element_text(hjust = 0.5, face = 'bold'), legend.position = c(0.9, 0.3),
        legend.background = element_blank(), legend.box.background = element_rect(color = 'black')) +
  coord_flip() +
  ggtitle('Payment methods in Farmer Markets') +
  scale_fill_manual(values = c('#EB7070', '#FEC771', '#E6E56C', '#64E291'))
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```



Credit card is the most common way of transaction offered in the farmers market.

Removing the variables

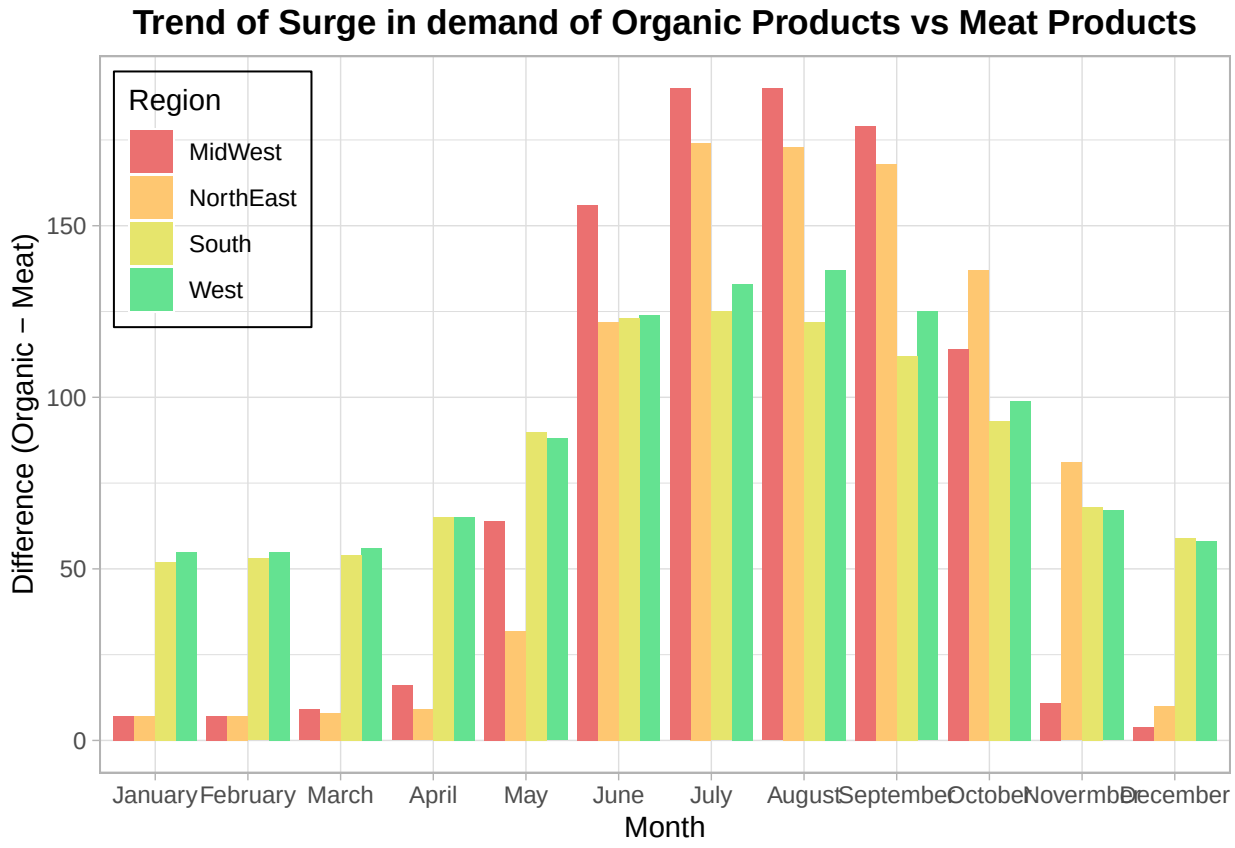
```
rm(credit_count, sfmnp_count, snap_count, task3_df, wic_count, wiccash_count)
```

Analysis of the meat and organice food consumption

```
meat_df = dataset %>%
  group_by(Region) %>%
  mutate(Category = ifelse(Eggs == 'Y' | Seafood == 'Y' |
                             Meat == 'Y' | Poultry == 'Y' | Tofu == 'Y', 'Meat Product', ''))
meat_df = meat_df %>%
  filter(Category == 'Meat Product' & Season1Date != '') %>%
  select(Category, Season1Date, Region)

meat_df$StartDate = str_split(meat_df$Season1Date, ' to ', simplify = TRUE)[, 1]
meat_df$EndDate = str_split(meat_df$Season1Date, ' to ', simplify = TRUE)[, 2]

meat_df = meat_df %>%
  filter(!is.na(mdy(StartDate)) & !is.na(mdy(EndDate)))

meat_df = meat_df %>%
  mutate(StartDate = month(mdy(StartDate))) %>%
  mutate(EndDate = month(mdy(EndDate)))
```

```
meat_df = meat_df %>%
  group_by(StartDate, EndDate) %>%
  mutate(Months = ifelse(StartDate < EndDate,
                         paste(months[StartDate:EndDate], collapse = ','),
                         paste(months[-((StartDate-1):(EndDate+1))], collapse = ',')))

meat_df = separate_rows(meat_df, Months, sep = ',', convert = TRUE)

meat_df$Months = factor(meat_df$Months, levels = months, labels = months_text)

meat_grouped_df = meat_df %>%
  group_by(Region, Months)%>%
  summarise(Count = n())

# Organic

organic_df = dataset %>%
  group_by(Region) %>%
  mutate(Category = ifelse(Organic == 'Y' | Vegetables == 'Y' | Beans =='Y' |
                             WildHarvested == 'Y' | Mushrooms == 'Y','Organic Product', ''))
organic_df = organic_df %>%
  filter(Category == 'Organic Product' & Season1Date != '') %>%
  select(Category, Season1Date, Region)

organic_df$StartDate = str_split(organic_df$Season1Date, ' to ', simplify = TRUE)[, 1]
organic_df$EndDate = str_split(organic_df$Season1Date, ' to ', simplify = TRUE)[, 2]

organic_df = organic_df %>%
  filter(!is.na(mdy(StartDate)) & !is.na(mdy(EndDate)))

organic_df = organic_df %>%
  mutate(StartDate = month(mdy(StartDate))) %>%
  mutate(EndDate = month(mdy(EndDate)))

organic_df = organic_df %>%
  group_by(StartDate, EndDate) %>%
  mutate(Months = ifelse(StartDate < EndDate,
                         paste(months[StartDate:EndDate], collapse = ','),
                         paste(months[-((StartDate-1):(EndDate+1))], collapse = ',')))

organic_df = separate_rows(organic_df, Months, sep = ',', convert = TRUE)

organic_df$Months = factor(organic_df$Months, levels = months, labels = months_text)

organic_grouped_df = organic_df %>%
  group_by(Region, Months)%>%
  summarise(Count = n())

meat_organic = rbind(meat_grouped_df, organic_grouped_df)
meat_organic_percent = meat_organic %>%
  group_by(Region, Months)%>%
  mutate(Difference = diff(Count))
```

```
ggplot(data = meat_organic_percent) +
  geom_bar(aes(x = Months, y = Difference, fill = Region), stat = 'identity', position = 'dodge') +
  theme_light() +
  xlab('Month') +
  ylab('Difference (Organic - Meat)') +
  ggtitle('Trend of Surge in demand of Organic Products vs Meat Products') +
  theme(plot.title = element_text(hjust = 0.5, face = 'bold'), legend.position = c(0.1, 0.8),
        legend.background = element_blank(), legend.box.background = element_rect(color = 'black'))+
  scale_fill_manual(values = c('#EB7070', '#FEC771', '#E6E56C', '#64E291'))
```



In order to analyse this plot, we'll have to consider the findings that we have obtained till now. Northeast has very less number of markets, still has almost the same peak for th bar as Midwest which has the second highest number of farmer markets. The height of the bar represents the difference between the organic product and meat product, meaning the markets in Northeast offer more organic food than meat products.

Removing the variables.

```
rm(meat_df, meat_grouped_df, meat_organic, meat_organic_percent, organic_df, organic_grouped_df, months
```